WEST

Help | Logout

Main Menu | Search Form | Posting Counts | Show S Numbers | Edit S Numbers

Search Results -

Terms	Documents
l10 and (protein or dna)	31

Database: US Patents Full-Text Database

Refine Search: l10 and (protein or dna)

Search History

DB Name	Query	Hit Count	Set Name
USPT	l10 and (protein or dna)	31	L11
USPT	(three-dimension$) with database$	181	L10
USPT	(three-dimension$) and database$	2034	L9
USPT	l3 and bioinfor$	2	L8
USPT	l5 and bioinfor$	0	L7
USPT	l5 and (structure$ with database)	42	L6
USPT	l4 and (crystal structure$)	119	L5
USPT	l3 and x-ray	220	L4
USPT	l2 and database$	584	L3
USPT	(three-dimensio$) and protein	4038	L2
USPT	pan-geneomic	0	L1

(FILE 'HOME' ENTERED AT 16:00:38 ON 28 SEP 1999)

FILE 'BIOSIS, CAPLUS, MEDLINE, SCISEARCH, COMPUSCIENCE, COMPUAB' ENTERED AT 16:01:06 ON 28 SEP 1999

L1 18458 S DATABASE (P) SEQUENCE
L2 5900 S L1 AND STRUCT?
L3 751 S L2 AND GENOMIC
L4 26 S L3 AND BIOINFORMATIC?
L5 18 DUP REM L4 (8 DUPLICATES REMOVED)

09/235,986

=> d l5 bib,ab 1-18

L5 ANSWER 1 OF 18 SCISEARCH COPYRIGHT 1999 ISI (R)
AN 1999:456298 SCISEARCH
GA The Genuine Article (R) Number: 204CG
TI **Genomics**, mutations and the Internet: The naming and use of parts
AU Scriver C R (Reprint); Nowacki P M
CS MCGILL UNIV, MONTREAL CHILDRENS HOSP, RES INST, DEBELLE LAB, 2300 TUPPER ST, MONTREAL, PQ H3H 1P3, CANADA (Reprint); MCGILL UNIV, DEPT BIOL, MONTREAL, PQ, CANADA; MCGILL UNIV, DEPT PEDIAT, MONTREAL, PQ H3A 2T5, CANADA; MCGILL UNIV, DEPT HUMAN GENET, MONTREAL, PQ, CANADA
CYA CANADA
SO JOURNAL OF INHERITED METABOLIC DISEASE, (MAY 1999) Vol. 22, No. 4, pp. 519-530.
Publisher: KLUWER ACADEMIC PUBL, SPUIBOULEVARD 50, PO BOX 17, 3300 AA DORDRECHT, NETHERLANDS.
ISSN: 0141-8955.
DT Article; Journal
FS LIFE
LA English
REC Reference Count: 54
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS
AB Mutations are the source of genetic variation and diversity; by their effect, some are neutral, others are pathogenic. In contemporary genetics, mutations appear at the interface between **genomics** (**structural** and functional) and genetics (heredity), where they serve gene discovery and mapping (**genomics**) and generate challenges to modify their phenotypic effects (medical genetics). Assuming the human genome harbours 80000 transcribed genes each possessing at least 100 different (germline) alleles in a typical population, how then to record and recover data on at least 8 million human alleles? **Bioinformatics** is the essential resource to create the corresponding accessible digital libraries (**genomic** and locus-specific mutation databases) for this purpose, a goal to which The HUGO Mutation Database Initiative (Science 279: 10-11, 1998) aspires. Guidelines now exist for naming alleles (Hum Mutat 11: 1-3, 1998). The principles behind the practice are illustrated by PAHdb (http://www.mcgill.ca/pahdb), a prototype locus-specific mutation database (NAR 26: 220-225, 1998), and by prototype **genomic** mutation databases (HGMD (NAR 26: 285-287, 1998), http://www.uwcm.ac.uk/uwcm/mg/hgmd0.html; the EBI mutation database, http://www2.ebi.ac.uk/mutations/; and OMIM, http://www.ncbi.nlm.nih.gov/Omim.html).

L5 ANSWER 2 OF 18 CAPLUS COPYRIGHT 1999 ACS
AN 1999:146268 CAPLUS
TI Protein NMR and the human proteome project
AU Montelione, Gaetano T.
CS Center for Advanced Biotechnology Medicine and Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, NJ, 08854, USA
SO Book of Abstracts, 217th ACS National Meeting, Anaheim, Calif., March 21-25 (1999), POLY-232 Publisher: American Chemical Society, Washington, D. C.
CODEN: 67GHA6
DT Conference; Meeting Abstract
LA English

AB Genome sequencing projects are rapidly identifying all the genes in several organisms. The products of these genes are widely recognized as the next generation of therapeutics and targets for the development of pharmaceuticals. While identification of these genes is proceeding quickly, elucidation of their three dimensional (3D) **structures** and biochem. functions lags far behind. In some cases, knowledge of 3D **structures** of proteins can provide important insights into **structural** homol. that is not easily recognized by **sequence** alignment comparisons. Thus, anal. of a protein's 3D **structure** by NMR or X-ray crystallog. prior to characterization of the protein's biochem. function can sometimes provide key information regarding protein fold class, locations and clustering of conserved residues, and surface electrostatic field distributions. This information can be used to develop hypotheses regarding potential biochem. functions, and the resulting limited set of putative biochem. functions tested by appropriate biochem. assays. NMR chem. shift assignments and soln. **structures** of proteins also provide the basis for epitope-mapping, mol. dynamics, and SAR studies, and set the stage for subsequent drug development using combinatorial and/or rational design methods. We are developing technologies that will significantly accelerate the process of **structure** detn. by NMR. These include **bioinformatics** methods for parsing novel genes into domain encoding regions, high-level "multiplexed" protein expression systems, and NMR pulse **sequences**, data collection methods, and expert-system software for automated anal. of protein resonance assignments and 3D **structures**. These technologies and the resulting exptl. data are being organized and integrated using relational **databases**. The goal of this work is to develop a "high-throughput" process for **structural** anal. of novel gene products on a **genomic** scale. In a pilot project, these techniques are being applied to clusters of orthologous genes coding for proteins of unknown **structure** and function, with the aim of testing the hypothesis that 3D **structural** anal. can sometimes provide useful and important clues regarding the biochem. functions of orphan gene products. The relationship of our effort and the emerging international interest in a large-scale Human Proteome Project will be discussed.

L5 ANSWER 3 OF 18 BIOSIS COPYRIGHT 1999 BIOSIS
AN 1999:228146 BIOSIS
DN PREV199900228146
TI FORESST: Fold recognition from secondary **structure** predictions of proteins.
AU Di Francesco, V. (1); Munson, P. J. (1); Garnier, J. (1)
CS (1) Analytical Biostatistics Section, Mathematical and Statistical Computing Laboratory, Institute, Center for Information Technology, National Institutes of Health, Bethesda, MD, 20892-5626 USA
SO Bioinformatics (Oxford), (Feb., 1999) Vol. 15, No. 2, pp. 131-140. ISSN: 1367-4803.
DT Article
LA English
SL English
AB Motivation: A method for recognizing the three-dimensional fold from the protein amino acid **sequence** based on a combination of hidden Markov models (HMMs) and secondary **structure** prediction was recently developed for proteins in the Mainly-Alpha **structural** class. Here, this methodology is extended to Mainly-Beta and Alpha-Beta class proteins. Compared to other fold recognition methods based on HMMs, this approach is novel in that only secondary **structure** information is used. Each HMM is trained from known secondary **structure sequences** of proteins having a similar fold. Secondary **structure** prediction is performed for the amino acid **sequence** of a query protein. The predicted fold of a query protein is the fold described by the model fitting the predicted **sequence** the best. Results: After model cross-validation, the success rate on 44 test proteins covering the three **structural** classes was found to be 59%. On seven fold predictions performed prior to the publication of experimental **structure**, the success rate was 71%. In conclusion, this approach manages to capture important information about the fold of a protein embedded in the length and arrangement of the predicted helices,

strands and coils along the polypeptide chain. When a more extensive library of HMMs representing the universe of known **structural** families is available (work in progress), the program will allow rapid screening of **genomic databases** and **sequence** annotation when fold similarity is not detectable from the amino acid **sequence**. Availability: FORESST web server at http://absalpha.dcrt.nih.gov:8008/ for the library of HMMs of **structural** families used in this paper. FORESST web server at http://www.tigr.org/ for a more extensive library of HMMs (work in progress). Contact: valedf@tigr.org; munson@helix.nih.gov; garnier@helix.nih.gov.

L5 ANSWER 4 OF 18 CAPLUS COPYRIGHT 1999 ACS
AN 1999:94567 CAPLUS
DN 130:266967
TI The PIR-international protein **sequence database**
AU Barker, Winona C.; Garavelli, John S.; McGarvey, Peter B.; Marzec, Christopher R.; Orcutt, Bruce C.; Srinivasarao, Geetha Y.; Yeh, Lai-Su L.; Ledley, Robert S.; Mewes, Hans-Werner; Pfeiffer, Friedhelm; Tsugita, Akira; Wu, Cathy
CS Protein Information Resource, National Biomedical Research Foundation, Washington, DC, 20007, USA
SO Nucleic Acids Res. (1999), 27(1), 39-43
CODEN: NARHAD; ISSN: 0305-1048
PB Oxford University Press
DT Journal
LA English
AB The Protein Information Resource (PIR; http://www-nbrf.georgetown.edu/pir/) supports research on mol. evolution, functional **genomics**, and computational biol. by maintaining a comprehensive, non-redundant, well-organized and freely available protein **sequence database**. Since 1988 the **database** has been maintained collaboratively by PIR-International, an international assocn. of data collection centers cooperating to develop this resource during a period of explosive growth in new **sequence** data and new computer technologies. The PIR Protein **Sequence Database** entries are classified into superfamilies, families and homol. domains, for which **sequence** alignments are available. Full-scale family classification supports comparative **genomics** research, aids **sequence** annotation, assists **database** organization and improves **database** integrity. The PIR WWW server supports direct online **sequence** similarity searches, information retrieval, and knowledge discovery by providing the Protein **Sequence Database** and other supplementary **databases**. **Sequence** entries are extensively cross-referenced and hypertext-linked to major nucleic acid, literature, genome, **structure**, **sequence** alignment and family **databases**. The weekly release of the Protein **Sequence Database** can be accessed through the PIR Web site. The quarterly release of the **database** is freely available from our anonymous FTP server and is also available on CD-ROM with the accompanying ATLAS **database** search program.

L5 ANSWER 5 OF 18 CAPLUS COPYRIGHT 1999 ACS
AN 1998:481365 CAPLUS
DN 129:197468
TI **Genomics** and drugs: finding the optimal drug for the right patient
AU Sadee, Wolfgang
CS Biopharmaceutical Sciences and Pharmaceutical Chemistry, UCSF, San Francisco, CA, USA
SO Pharm. Res. (1998), 15(7), 959-963
CODEN: PHREEB; ISSN: 0724-8741
PB Plenum Publishing Corp.
DT Journal; General Review
LA English
AB A review with 32 refs. With an exponential growth of the **sequence database**, novel approaches are needed to benefit from this vast newly available information. This is the goal of **bioinformatics**

, the science dealing with the management and integrati of information on **sequence**, **structure** and function. Within a few years, **genomics** and **bioinformatics** have taken centerstage in the biosciences, and any pharmaceutical company of repute is establishing a strong effort in this area. The author attempts to analyze what **genomics** might mean for drug discovery, development and clin. application. How academia and industry have to adapt to the challenge of providing properly trained scientists and reorienting global research directions is implicit in this discussion.

L5 ANSWER 6 OF 18 CAPLUS COPYRIGHT 1999 ACS
AN 1998:318728 CAPLUS
DN 129:108681
TI From sequence to **structure** to literature: the protocol approach to bioinformation
AU Wu, O. P.; Seow, K. T.; Wong, L.; Chung, S. Y.; Subbiah, S.
CS BioInformatics Center, National University of Singapore, Singapore, 119260, Singapore
SO Pac. Symp. Biocomput. '98 (1998), 747-758. Editor(s): Altman, Russ B. Publisher: World Scientific, Singapore, Singapore. CODEN: 66CDAZ
DT Conference
LA English
AB Until the recent advent of high-throughput exptl. data-acquisition in biol., the computational anal. of the biol. data was predominantly on an ad hoc basis - i.e. the application of a given piece of software on the biol. data depended on the need of the moment. This "functional approach" often resulted in piecemeal computational anal. with large amt. of intervening "dead-time". The present high-throughput availability of exptl. biol. data requires a more streamlined and integrated "protocol approach". In this work, we illustrate such a user-friendly protocol using a common question frequently faced by a wet-lab bench-biologist - "Now that I have a DNA or protein sequence, what can I do with it using a computer". As phrased, this question is steeped in the functional approach. In contrast, the protocol approach would re-phrase the same question as "Now that I have a DNA or protein sequence, what can a computer do for me". Our integrating tool can start with a sequence and build a substantial custom data-warehouse of computationally derived sequence information, **structure** information and relevant published literature, that is continually updated.

L5 ANSWER 7 OF 18 CAPLUS COPYRIGHT 1999 ACS
AN 1998:318726 CAPLUS
DN 129:108680
TI Development of software tools at **Bioinformatics** Center (BIC) at the National University of Singapore (NUS)
AU Kolatkar, P. R.; Sakharkar, M.; Tse, C. R.; Kiong, B. K.; Wong, L.; Tan, T. W.; Subbiah, S.
CS BioInformatics Centre, National University of Singapore, Singapore, Singapore
SO Pac. Symp. Biocomput. '98 (1998), 735-746. Editor(s): Altman, Russ B. Publisher: World Scientific, Singapore, Singapore. CODEN: 66CDAZ
DT Conference
LA English
AB There is a burgeoning vol. of information and data arising from the rapid research and unprecedented progress in mol. biol. This has been particularly affected by the Human Genome Project which is trying to completely **sequence** three billion nucleotides of the human genome (1),(1a). Other genome sequencing projects are also contributing substantially to this exponential growth in the no. of DNA nucleotides and proteins sequenced. The no. of journals, reports and research papers and tools required for the anal. of these **sequences** has also increased. For this the life sciences today needs tools in information technol. and computation to prevent degeneration of this data into an inchoate accretion of unconnected facts and figures. The recently formed **BioInformatics** Center (BIC) at the National University of Singapore (NUS) provides access to various commonly used computational tools available over the World Wide Web (WWW) - using a uniform interface

and easy access. We have also come up with a new **database** tool, BioKleisli, which allows you to interact with various [redacted]. scattered, heterogeneous, **structurally** complex and constantly evolving data sources. This paper summarizes the importance of network access and **database** integration to biomedical research and gives a glimpse of current research conducted at BIC.

L5 ANSWER 8 OF 18 CAPLUS COPYRIGHT 1999 ACS
AN 1999:175458 CAPLUS
DN 131:68903
TI A study using sequence comparison to investigate the molecular evolution of mitochondrial tRNA genes
AU Sagara, Jun-Ichi; Nakamura, Shugo; Ikeguchi, Mitsunori; Shimizu, Kentaro
CS Department of Biotechnology, The University of Tokyo, Tokyo, 113-8657, Japan
SO Genome Inf. Ser. (1998), 9, 353-354
CODEN: GINSE9; ISSN: 0919-9454
PB Universal Academy Press
DT Journal
LA English
AB The authors developed a computational method based on principal component anal. (PCA) and multidimensional scaling anal. (MDS), and in the present study have used it to detect bases characterizing specific **sequences** of mitochondrial tRNA genes. The authors' methods first classify the **sequence** in a **genomic database** into groups by PCA of multiple **sequence** alignment: Gene **sequences** are represented as vectors in a generalized **sequence** space, and groups of similar **sequences** are revealed when these vectors are projected onto a lower-dimensional **sequence** space. The distribution of bases is then compared with the distribution of **sequences** by using MDS, in which the bases of each **sequence** are projected individually onto the same **sequence** space. This makes it possible to identify bases characteristic of each group. In evaluating their method, they used full-length mitochondrial tRNA gene **sequences** deposited in the aligned **sequence database** of European **Bioinformatics** Institute (EBI) . The authors detected many bases characteristic of all the mitochondrial tRNA gene **sequences** of various species. Most of the characteristic bases are in stem regions, and most of the characteristic bases that are not in stem regions are in T and D domains, which are elbow regions of tRNAs. The results suggest that the characteristic bases in these stems and domains have a role of preserving the L-shape **structure** of each tRNA.

L5 ANSWER 9 OF 18 BIOSIS COPYRIGHT 1999 BIOSIS
AN 1998:346243 BIOSIS
DN PREV199800346243
TI GeneGenerator: A flexible algorithm for gene prediction and its application to maize sequences.
AU Kleffe, Juergen (1); Hermann, Klaus (1); Vahrson, Wolfgang (1); Wittig, Burghardt (1); Brendel, Volker
CS (1) Freie Univ. Berlin, Abt. Mol. Bioinformatik, Inst. Molekularbiol. Biochem., Arnimallee 22, 14195 Berlin Germany
SO Bioinformatics (Oxford), (1998) Vol. 14, No. 3, pp. 232-243.
ISSN: 1367-4803.
DT Article
LA English
AB Motivation: We developed GeneGenerator because of the need for a tool to predict gene **structure** without knowing in advance how to score potential exons and introns in order to obtain the best results, pertinent in particular to less well-studied organisms for which suitable training sets are small. GeneGenerator is a very flexible algorithm which for a given **genomic sequence** generates a number of feasible gene **structures** satisfying user-defined constraints. The specific implementation described in detail requires minimum scoring for translation start and donor and acceptor splice sites according to previously trained logitlinear models. In addition, potential exons and introns are required to exceed specified minimal lengths and threshold scores for coding or non-coding potential derived as log-likelihood ratios

of appropriate Markov sequence models. Results: A database of 46 non-redundant genomic sequences from maize is used for illustration. It is shown that the correct gene structures do not always maximize the considered target function. However, in most cases, the correct or nearly correct structures are found in a small set of high-scoring structures. A critical review of the generated structures sometimes allows the choices to be narrowed by considering additional variables such as predicted splice site strength or local optimality of splice site scores. Summary statistics for prediction accuracy over all 46 maize genes are derived under cross-validation and non-cross-validation training conditions for the Markov sequence models. The algorithm achieved exon sensitivity of 0.81 and specificity of 0.75 on an independent set of 14 novel maize genomic segments. Availability: GeneGenerator runs under Borland-Pascal 7.0 using MS-DOS and C on UNIX work stations. The source code is available upon request. Contact: jkleffe@euler.grumed.fu-berlin-de

L5 ANSWER 10 OF 18 CAPLUS COPYRIGHT 1999 ACS
AN 1998:318653 CAPLUS
DN 129:104984
TI An editing environment for DNA sequence analysis and annotation
AU Uberbacher, Edward C.; Xu, Ying; Shah, Manesh B.; Olman, Victor; Parang, Morey; Mural, Richard J.
CS Computer Science and Mathematics Divisions, Oak Ridge National Laboratory, Oak Ridge, TN, 37831-6364, USA
SO Pac. Symp. Biocomput. '98 (1998), 217-227. Editor(s): Altman, Russ B. Publisher: World Scientific, Singapore, Singapore. CODEN: 66CDAZ
DT Conference
LA English
AB This paper presents a computer system for analyzing and annotating large-scale genomic sequences. The core of the system is a multiple-gene structure identification program, which predicts the most "probable" gene structures based on the given evidence, including pattern recognition, EST and protein homol. information. A graphics-based user interface provides an environment which allows the user to interactively control the evidence to be used in the gene identification process. To overcome the computational bottleneck in the database similarity search used in the gene identification process, we have developed an effective way to partition a database into a set of sub-databases of "related" sequences, and reduced the search problem on a large database to a signature identification problem and a search problem on a much smaller sub-database. This reduces the no. of sequences to be searched from N to O(.sqroot.N) on av., and hence greatly reduces the search time, where N is the no. of sequences in the original database. The system provides the user with the ability to facilitate and modify the anal. and modeling in real time.

L5 ANSWER 11 OF 18 CAPLUS COPYRIGHT 1999 ACS DUPLICATE 1
AN 1998:785231 CAPLUS
DN 130:121579
TI Mass spectrometric identification and microcharacterization of proteins from electrophoretic gels: strategies and applications
AU Jensen, Ole Norregaard; Larsen, Martin R.; Roepstorff, Peter
CS Department of Molecular Biology, Odense University, Odense, Den.
SO Proteins: Struct., Funct., Genet. (1998), (Suppl. 2), 74-89 CODEN: PSFGEY; ISSN: 0887-3585
PB Wiley-Liss, Inc.
DT Journal; General Review
LA English
AB A review with 77 refs. including the authors' own research. The entire genomic DNA sequences of a no. of prokaryotic and eukaryotic species are now available and many more, including the human genome, will be completed in the near future. The state-of-life of a cell at any given time, however, is defined by its protein compn., i.e., its proteome. Gel electrophoresis, mass spectrometry, and bioinformatics will be important tools for protein and proteome

anal. in the post-genome era. Protein identification from electrophoretic gels by mass spectrometric peptide mapping or peptide sequencing combined with **sequence database** searching is established and has been applied to numerous biol. systems. We describe current strategies and selected applications in mol. and cell biol. The next challenges are detailed **structure**/function analyses, which include studying the mol. compn. of multiprotein complexes and characterization of secondary modifications of proteins. The advantages and limitations of a no. of mass spectrometry-based strategies designed for microcharacterization of low amts. of protein from electrophoretic gels are discussed and illustrated by examples.

L5 ANSWER 12 OF 18 BIOSIS COPYRIGHT 1999 BIOSIS DUPLICATE 2
AN 1998:271604 BIOSIS
DN PREV199800271604
TI Leveraging **genomics** for the discovery of diagnostic markers.
AU Burczak, John D. (1); Cafferkey, Robert; Wilkinson, Francis E.
CS (1) diaDexus R and D, 3303 Octavius Drive, Santa Clara, CA 95054 USA
SO Journal of Clinical Ligand Assay, (Spring, 1998) Vol. 21, No. 1, pp. 47-57.
ISSN: 1081-1672.
DT General Review
LA English
AB **Genomics** is the study of all the genes in an organism, including their **sequence**, **structure**, regulation, interaction and products. Because nearly all diseases have a genetic component, **genomics** is an important tool for diagnostic medicine, both traditional biochemical testing as well as genetic testing. cDNA/expression (**genomic**) **databases** and **bioinformatics** have emerged as valuable tools within **genomic** research. These tools have already aided in the discovery of genes related to colorectal cancer, atherosclerosis, osteoporosis and a variety of other diseases. Gene **sequence** information yields insight into the identity and function of unknown genes, while comparison of gene expression patterns among normal and diseased tissues helps define the role of specific genes in a disease. The concomitant development of new gene analysis technologies, such as DNA arrays for rapid screening of multiple genes, is enhancing the discovery and evaluation of potential diagnostic markers. These technologies allow multiple diagnostic leads to be evaluated at once, thus achieving an economy of scale, as well as allow analysis of complex gene interactions. Several challenges exist for **genomics** in the area of diagnostics, including the discovery and analysis of low level expression genes and the development of versatile genetic testing instrumentation. As these challenges are overcome, the impact of **genomics** on diagnostic medicine will increase.

L5 ANSWER 13 OF 18 BIOSIS COPYRIGHT 1999 BIOSIS DUPLICATE 3
AN 1997:453491 BIOSIS
DN PREV199799752694
TI Drosophila-related expressed sequences.
AU Banfi, Sandro; Borsani, Giusepep; Bulfone, Alessandro; Ballabio, Andrea (1)
CS (1) Telethon Inst. Genetics Med., San Raffaele Biomedical Science Park, via Olgettina 58, Milan 20132 Italy
SO Human Molecular Genetics, (1997) Vol. 6, No. 10, pp. 1745-1753.
ISSN: 0964-6906.
DT General Review
LA English
AB The study of model organisms has been instrumental towards the elucidation of the basic mechanisms of human biology. Drosophila melanogaster has been the target of extensive genetic analyses over the past 90 years and a notable amount of information is known about its gene **structure**, gene regulation and gene function. The vast gene resource generated by the expressed sequence tags (ESTs) efforts was exploited to identify, using a **bioinformatic** approach, novel human and murine gene transcripts homologous to Drosophila mutant genes. A systematic characterization of these genes, named Drosophila-related expressed sequences (DRES), was performed including **genomic** mapping in human and mouse and detailed study of their expression pattern by RNA in situ hybridization

experiments. Comparison between DRES genes and their putative partners in Drosophila contributes to the understanding of their function in mammals and to the discovery of their possible role in disease.

L5 ANSWER 14 OF 18 CAPLUS COPYRIGHT 1999 ACS DUPLICATE 4
AN 1997:448565 CAPLUS
DN 127:117877
TI Computational gene discovery and human disease
AU Rawlings, Christopher J.; Searls, David B.
CS Dep. Bioformatics, SmithKline Beecham Pharmaceuticals, Harlow/Essex, CM19 5AW, UK
SO Curr. Opin. Genet. Dev. (1997), 7(3), 416-423
CODEN: COGDET; ISSN: 0959-437X
PB Current Biology
DT Journal; General Review
LA English
AB A review with 76 refs. **Bioinformatics** is now an essential tool in many aspects of human mol. genetics research. Methods for the prediction of gene **structure** are essential components in **genomic** sequencing projects and provide the key to deriving protein **sequence** and locating intron/exon junctions. **Sequence** comparison and data base searching are the pre-eminent approaches for predicting the likely biochem. function of new genes, although **sequence** profiles derived from families of aligned **sequences** have advantages in the detection of remote **sequence** relationships. The use of **sequence database** anal. for large-scale comparative anal. of genome **sequence** data from model organisms is emerging as the most important recent development in the application of **bioinformatics** methods for characterizing candidate disease genes.

L5 ANSWER 15 OF 18 CAPLUS COPYRIGHT 1999 ACS
AN 1998:469844 CAPLUS
DN 129:226597
TI Inferring gene **structures** in **genomic** sequences using pattern recognition and expressed sequence tags
AU Xu, Ying; Mural, Richard J.; Uberbacher, Edward C.
CS Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN, 37831-6364, USA
SO Proc. Int. Conf. Intell. Syst. Mol. Biol., 5th (1997), 344-353.
Editor(s): Gaasterland, Terry. Publisher: AAAI Press, Menlo Park, Calif.
CODEN: 66LJAU
DT Conference
LA English
AB Computational methods for gene identification in **genomic sequences** typically have two phases: coding region prediction and gene parsing. While there are many effective methods for predicting coding regions (exons), parsing the predicted exons into proper gene **structures**, to a large extent, remains an unsolved problem. This paper presents an algorithm for inferring gene **structures** from predicted exon candidates, based on Expressed **Sequence** Tags (ESTs) and biol. intuition/rules. The algorithm first finds all the related ESTs in the EST **database** (dbEST) for each predicted exon, and infers the boundaries of one or a series of genes based on the available EST information and biol. rules. Then it constructs gene models within each pair of gene boundaries, that are most consistent with the EST information. By exploiting EST information and biol. rules, the algorithm can (1) model complicated multiple gene **structures**, including embedded genes, (2) identify falsely-predicted exons and locate missed exons, and (3) make more accurate exon boundary predictions. The algorithm has been implemented and tested on long **genomic sequences** with a no. of genes. Test results show that very accurate (predicted) gene models can be expected when related ESTs exist for the predicted exons.

L5 ANSWER 16 OF 18 CAPLUS COPYRIGHT 1999 ACS
AN 1998:469836 CAPLUS
DN 129:226596
TI The gene-finder computer tools for analysis of human and model organisms

genome sequences
AU Solovyev, Victor; Sal[redacted], Asaf
CS Department of Cell Biology, Baylor College of Medicine, Houston, TX, 77030, USA
SO Proc. Int. Conf. Intell. Syst. Mol. Biol., 5th (1997), 294-302. Editor(s): Gaasterland, Terry. Publisher: AAAI Press, Menlo Park, Calif. CODEN: 66LJAU
DT Conference
LA English
AB We present a complex of new programs for promoter, 3'-processing, splice sites, coding exons and gene **structure** identification in **genomic** DNA of several model species. The human gene **structure** prediction program FGENEH, exon prediction - FEXH and splice site prediction - HSPL have been modified for **sequence** anal. of Drosophila (FGENED, FEXD and DSPL), Caenorhabditis elegans (FGENEN, FEXN and NSPL), Yeast (FEXY and YSPL) and Plant (FGENEA, FEXA and ASPL) **genomic sequences**. We recomputed all frequency and discriminant function parameters for these organisms and adjusted organism specific minimal intron lengths. An accuracy of coding region prediction for these programs is similar with the obsd. accuracy of FEXH and FGENEH. We have developed FEXHB and FGENEHB programs combining pattern recognition features and information about similarity of predicted exons with known **sequences** in protein **databases**. These programs have approx. 10% higher av. accuracy of coding region recognition. Two new programs for human promoter site prediction (TSSG and TSSW) have been developed which use Gosh (1993) and Wingender (1994) data bases of functional motifs, resp. POLYAH program was designed for prediction of 3'-processing regions in human genes and CDSB program was developed for bacterial gene prediction. We have developed a new approach to predict multiple genes based on double dynamic programming, that is very important for anal. of long **genomic** DNA fragments generated by genome sequencing projects. Anal. of uncharacterized **sequences** based on our methods is available through the University of Houston, Weizmann Institute of Science email servers and several Web pages at Baylor College of Medicine.

L5 ANSWER 17 OF 18 BIOSIS COPYRIGHT 1999 BIOSIS DUPLICATE 5
AN 1998:98258 BIOSIS
DN PREV199800098258
TI **Bioinformatics** in drug discovery.
AU Kingsbury, David T. (1)
CS (1) Chiron Corp., 4560 Horton St., Mailstop G350, Emeryville, CA 94608-2916 USA
SO Drug Development Research, (July-Aug., 1997) Vol. 41, No. 3-4, pp. 120-128.
ISSN: 0272-4391.
DT Article
LA English
AB It is estimated that when the Human Genome Project's DNA sequencing phase reaches it peak steady-state, the production rate will be approximately 70 to 80 "new" genes each day, every day. This estimate is based on **genomic** sequencing and not expressed **sequence** tags (ESTs) or full-length cDNA sequencing. The challenge to research scientists is to identify these "genes" to some level of detail and determine which are of potential value, and to exploit that information as quickly as possible. This is a daunting task even with powerful computer and information systems; however, without a sophisticated informatics infrastructure it is impossible. The supporting information infrastructure can be broken into a number of components: data-acquisition systems (including inventory control and reagent manipulation systems, **sequence** production software, etc.); data-analysis systems (including **sequence** analysis software, **structure** prediction software, gene mapping software, feature extraction software, etc.); and data-management system (including local and shared **databases**). **Databases** form the core systems for the identification of "new" genome features and are the core of any informatics-based drug development strategy. The development and appropriate use of **database** collections, often referred to as data warehouses, is an essential process and computer science has provided

tools for the extraction of knowledge from these databases collections. In high throughput research environments, automated systems for processing new data through these data warehouses are essential tools for discovery research.

L5 ANSWER 18 OF 18 SCISEARCH COPYRIGHT 1999 ISI (R)
AN 96:616933 SCISEARCH
GA The Genuine Article (R) Number: VC203
TI **BIOINFORMATICS** - PRINCIPLES AND POTENTIAL OF A NEW MULTIDISCIPLINARY TOOL
AU BENTON D, (Reprint)
CS NATL INST HLTH, NATL CTR HUMAN GENOME RES, 38 LIB DR, BETHESDA, MD, 20892 (Reprint)
CYA USA
SO TRENDS IN BIOTECHNOLOGY, (AUG 1996) Vol. 14, No. 8, pp. 261-272. ISSN: 0167-7799.
DT General Review; Journal
FS LIFE; AGRI
LA ENGLISH
REC Reference Count: 116
ABSTRACT IS AVAILABLE IN THE ALL AND IALL FORMATS
AB The materials of **bioinformatics** are biological data, and its methods are derived from a wide variety of computational techniques. Recent years have seen an explosive growth in biological data, and the development of novel computational methods. These methods have become essential to research progress in **structural** biology, **genomics**, **structure**-based drug design and molecular evolution. The development and maintenance of a robust infrastructure of biological data is of equal importance if biotechnology is to take maximum advantage of research advances in a wide variety of fields. While **bioinformatics** has already made important contributions, it faces significant challenges as it matures.